Introduction to machine learning

### What is machine learning?

In machine learning we generally predict something. For example, the stock will be profitable or not. Is it a genuine or spam email or even India will win or loose the match?

### How machine learning is different from prime number finding algorithms?

Lets take one machine learning usecase example say “stock price predictor”. And compare it with general ( non machine learning ) prime number program. In prime number algorithm we have a sound logic. Such as if number is passing some fixed criteria it is one hundred percent prime number.

In case of “stock price predictor”, we first decide which parameters affect the stock price, by considering parameter values we define a model. and once our model is ready, we choose input values for our parameter and then get some resultant value.

Analyse and parse Data → define model → input value → result

Let’s see with simplest example,

*we first decide which parameters affect the stock price,*

there can be numerous parameter I will take only 2 fictitious parameters. let’s say stock price depends on date and inflation rate.

*by considering parameter values we define a model*

As we decide date and inflation rate as our parameters, we will try to get the data, the data should show some kind of relation between parameter and what we want to predict.

Date and Stock price

{<date>, <stock price>}

{1,10;2,11;3,9;4,14;5,10;6,11;7,9;8,14;9,10;10,11;11,9;12,14;13,10;14,11;15,9;16,14;17,10;18,11;19,9;20,14;21,10;22,10;23,11;24,9;25,14;26,10;27,10;28,11;29,9;30,14}

Such when date is 25 the closing stock price is 14.

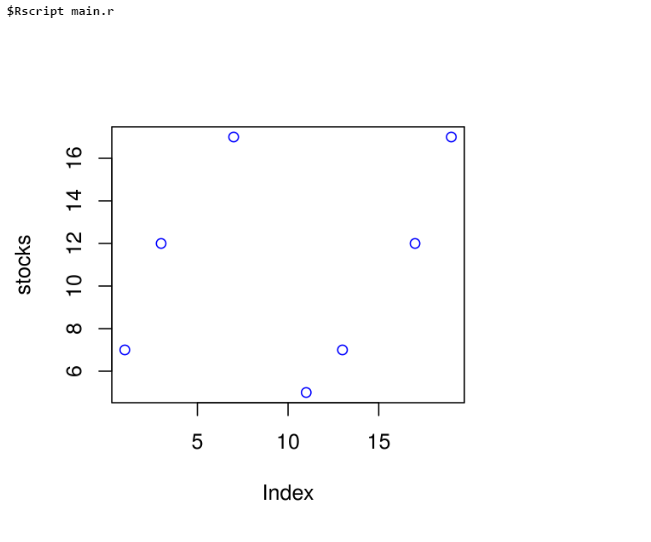
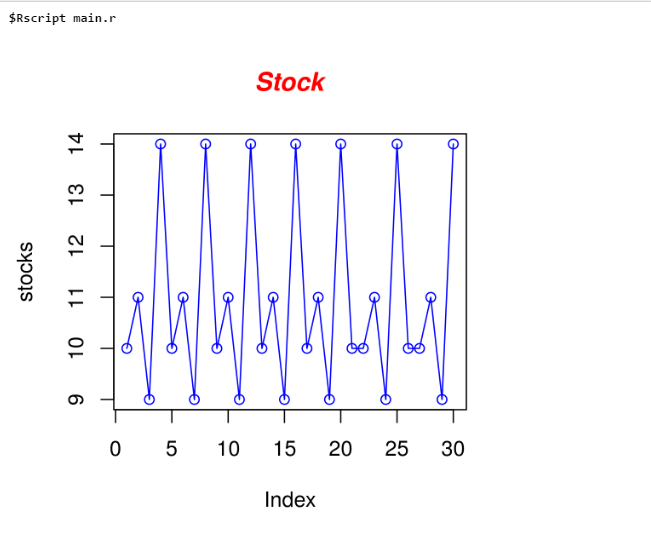
Inflation rate and stock price.

{inflation rate, stock price}

1,7;3,12;7,17;11,5;13,7;17,12;19,17

When inflation rate is 7 then closing stock price is 17.

Now we define a model, for time being consider model means plotting them on a graph.



As you can see there are two graphs, one which shows the relationship between date and stock and another one showing inflation rate and stock. The fun begins now,

We wanted to find out what is the stock price at 24 date when inflation is 12.

From graph one, the stock price will be 9

From graph two, the stock price will be 7

So defiantly the stock price can be 6,7,8,9,10

And might not be 1,2,3, 17,12,13,14

Now as I want some single value, from 7 and 9 I will do the mean of 7 and 9 which is 8.

Let’s find out another price, date: 3 and inflation is 7

So two values I got for stock 9 and 17, again I will find the mean, and my predicted stock price is 13.

Now something harder. Date: 4 but this time inflation is 2

We have a data for date but we don’t have data for inflation. but we do have data for inflation 1 and 3 which is 7 and 12. We will assume the data must be something around 9. (the average of 7 and 12).

With this information we will predict the stock price as 9.

Now we can use this computed value of inflation in our model. This is something we don’t do in our prime number type algorithms.

You might have a question in mind why I am doing mean and averages. The assumption is, the world follows continuity instead of too frequent ups and downs. Some of the greatest minds of human race developed many algorithms, one of the example is Linear regression. All are grouped under statistics. Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data (from wiki)

In machine learning,

we have an array of data, after initial parsing and cleaning, we form 2 parts of it (say A and B),

Then we “Train” the model, in this step, we basically form a model from one part of data (A).

In Training, we basically implement best suited Statistical algorithm on data.

After forming the model, we “Test” the model, with another part that do not took part in training the

model (B).

Finally, we request the prediction using actual input parameters.

As you can see the precision of resultant value is based on the data I have, the nature of such algorithm is different from the prime number finding algorithm.

Jumping on machine learning algorithm without having knowledge of Statistics is useless. One should at least know basics of statistics.

As you can see, the most important thing in deciding prediction is vast amount of data, if you have a good amount of data, you get better predictions. We stored this huge amount of data in Big Data Systems such as Hadoop, AWS or anything that is fast enough to handle the required data. Programming in Scala and Python is much easier than writing code in Java. We really have to write less code, thanks to filter, map and reduce functionality. I like python because in NLP almost all the time I am dealing with text data. I don’t need to take care of type casting.

Tips for blog

Focus on your audience!

Tip 1: Clear paragraphs

Tip 2: Short sentences

Tip 3: Limit difficult words

Tip 4: Use transition words

Tip 5: Mix it up!

Focus on using nouns and verbs; use adverbs and adjectives sparingly

Use active rather than passive verbs

5. Show, don’t tell

This is the best dog toy money can buy.

Or

We made the “Rough Rover” dog toy from durable, 100 percent natural rubber, designed to resist punctures and tears from even the most dedicated of chewers.

Which version gives you a clearer picture of the type of toy you’re buying? Specific details in the second description show readers the dog bone rather than tell them about it.

Make text scannable

use bulleted or numerical lists

Incorporate multimedia

11. Leave them wanting more